REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE September 2001	3. REPORT TYPE AND DATES COVERED Technical Report September 2001

4. TITLE AND SUBTITLE Empirical Bayes Estimation with Kernel Sequence Method	5. FUNDING NUMBERS DAAD19-00-1-0502
6. AUTHOR(S) Shanti S. Gupta & Jinjun Lu	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Purdue University	8. PERFORMING ORGANIZATION REPORT NUMBER Technical Report #01-11
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U. S. Army Research Office P.O. Box 12211 Research Triangle Park, NC 27709-2211	10. SPONSORING / MONITORING AGENCY REPORT NUMBER 40940.6-MA

11. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12 a. DISTRIBUTION / AVAILABILITY STATEMENT	12 b. DISTRIBUTION CODE
Approved for public release; distribution unlimited.	

13. ABSTRACT (

Abstract: In this paper, we consider the empirical Bayes estimation in the exponential family. A minimax lower bound is derived. It is shown that the best possible rate of empirical Bayes estimators is $O(1/n)$ if θ is bounded. Then we turn to find an empirical Bayes estimator with a rate close to this lower bound rate. Applying the kernel sequence method, we are able to construct an empirical Bayes estimator with a rate of $O(n^{-1}(\ln n)^7(\ln\ln n)^2)$. Under the same assumption, this rate is the fastest compared to the earlier results published in the literature.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Empirical Bayes, Exponnentional family, kernel sequence method, regret, rate of convergence	22
	16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION ON THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev.2-89)
Prescribed by ANSI Std. 239-18
298-102

20011101 135

EMPIRICAL BAYES ESTIMATION WITH KERNEL SEQUENCE METHOD[1]

by

Shanti S. Gupta and Jianjun Li
Purdue University

Technical Report # 01-11

Department of Statistics
Purdue University
West Lafayette, IN USA

September 2001

Empirical Bayes Estimation with Kernel Sequence Method [1]

Shanti S. Gupta *and* Jianjun Li

Department of Statistics, Purdue University, W. Lafayette, IN 47907

Department of Statistics, Purdue University, W. Lafayette, IN 47907

Abstract: In this paper, we consider the empirical Bayes estimation in the exponential family. A minimax lower bound is derived. It is shown that the best possible rate of empirical Bayes estimators is $O(1/n)$ if θ is bounded. Then we turn to find an empirical Bayes estimator with a rate close to this lower bound rate. Applying the kernel sequence method, we are able to construct an empirical Bayes estimator with a rate of $O(n^{-1}(\ln n)^7(\ln\ln n)^2)$. Under the same assumption, this rate is the fastest compared to the earlier results published in the literature.

AMS 1991 Subject classification: Primary 62C12, 62G05.

Key words and phrases: Empirical Bayes; exponentional family; kernel sequence method; regret; rate of convergence.

[1] This research was supported in part by US Army Research Office, Grant DAAD19-00-1-0502 at Purdue University.

1 Introduction

Suppose that an observation X is obtained from a distribution with density

$$f(x|\theta) = c(\theta)\exp\{\theta x\}h(x), \qquad -\infty \le a < x < b \le +\infty, \tag{1}$$

where $h(x) > 0$ for $x \in (a,b)$ and $h(x)$ is bounded from below on any compact set of (a,b). The parameter θ is distributed according to an unknown and unspecified prior G on the parameter space Ω, a subset of the natural parameter space $\Omega_0 \equiv \{\theta : c(\theta) > 0\}$.

Suppose that one wants to estimate θ after observing $X = x$. Under the squared error loss, the Bayes estimator is given by $\phi_G(x) = E[\theta|X = x]$. It can be computed if G is known. In situations where G is unknown, this solution cannot work. A solution to these situations is to apply the empirical Bayes approach to construct an empirical Bayes estimator. This approach assumes that n independent past observations $X_1, \cdots, X_n$ are available. Thus an estimator of θ can be constructed based on $X_1, \cdots, X_n$ and $X = x$. The estimator is called the empirical Bayes estimator, and denoted by $\phi_n(x, X_1, \cdots, X_n) \equiv \phi_n(x) \equiv \phi_n$. The Bayes risk $R(\phi_G, G)$ is

$$R(\phi_G, G) = \int\int (\phi_G(x) - \theta)^2 f(x|\theta)\,dx\,dG(\theta). \tag{2}$$

The overall risk of ϕ_n, denoted by $R(\phi_n, G)$, is

$$R(\phi_n, G) = E[\int\int (\phi_n(x) - \theta)^2 f(x|\theta)\,dx\,dG(\theta)]. \tag{3}$$

$R(\phi_n, G) - R(\phi_G, G)$, the difference of the (overall) risk of ϕ_n and the Bayes risk, is call the regret of the estimator ϕ_n and used to measure the performance of ϕ_n.

The above estimation problem has been considered by many authors. (See Lin (1975), Singh (1976), Singh (1979), Pensky (1998) and the references listed there.) Singh (1979) significantly improved the previous results in terms of the rate of convergence. He constructed an empirical Bayes estimator and showed that the estimator has a rate of convergence of $O(-n^{\frac{(2r-1)+}{2r+1}})$. Pensky (1998) applied advanced wavelet techniques to construct empirical Bayes estimators and obtained a better rate.

For this empirical Bayes estimation problem, a natural question arises: what is the best possible rate ? To answer this question, a minimax lower bound of empirical Bayes estimators is derived and it is shown that the best possible rate is $O(1/n)$ if θ is distributed within a compact (bounded) set.

Also we shall construct an empirical Bayes estimator using kernel sequence method. The kernel sequence method enables us to use the C_∞-smoothness of $\alpha_G(x)$ and $\psi_G(x)$ (defined in Section 2). Thus improved estimators of $\alpha_G(x)$ and $\psi_G(x)$ are obtained. Based on these estimators, we construct an empirical Bayes estimator $\phi_n(x)$ and show that $\phi_n(x)$ has a rate of convergence of $O(n^{-1}(\ln n)^7(\ln\ln n)^2)$ under the assumption of $\Omega \subset [\theta_1, \theta_2] \subset \Omega_0$.

This paper is organized as follows: a minimax lower bound is derived in Section 2 by converting the global problem into a local problem, identifying the Bayes estimator as a functional of the marginal density of X, and constructing the hardest two-point subproblem. The construction of the estimator $\phi_n(x)$ is presented in Section 3 and its performance is also studied there. In Section 4, we present a few examples, which include three examples used in Singh (1979) and the comparisons of our results with his. The proofs are given in Section 5. In Section 6, we summarize our results and make some comparisons with the results published recently in the literature.

Finally, the readers may refer to Robbins (1956, 1964) to learn more about the empirical Bayes approach. As for applications of the empirical Bayes estimation, one may see Bendel and Carlin (1990), Louis (1991), Desouza (1991), Mollie and Richardson (1991), Norberg (1989), Lahiri and Park (1991), Chen and Singpurwalla (1996) and Pensky and Singh (1999).

2 Lower Bound of Empirical Bayes Estimators

We shall obtain a minimax lower bound for empirical Bayes estimators. This will show that the best possible rate for any empirical Bayes estimator is $O(1/n)$.

2.1 Conversion to a Local Problem

Under the squared error loss, the Bayes estimator $\phi_G(x)$ is the posterior mean of θ given $X = x$. Simple calculations show that

$$\phi_G(x) = E[\theta|X = x] = \frac{\int \theta c(\theta) \exp(\theta x) dG(\theta)}{\int c(\theta) \exp(\theta x) dG(\theta)}. \tag{4}$$

Let $\alpha_G(x) = \int c(\theta) \exp(\theta x) dG(\theta)$ and $\psi_G(x) = \int \theta c(\theta) \exp(\theta x) dG(\theta)$. Then the Bayes estimator of θ can be written as

$$\phi_G(x) = \frac{\psi_G(x)}{\alpha_G(x)}. \tag{5}$$

Suppose that the prior G has a compact support $[\theta_1, \theta_2]$ or its support belongs to the compact set $[\theta_1, \theta_2]$. Let $\mathcal{G}$ be the class of this type of priors, i.e.,

$$\mathcal{G} = \Big\{G : G \text{ has the support } \Omega \subset [\theta_1, \theta_2] \subset \Omega_0 = \{\theta : c(\theta) > 0\}\Big\}. \tag{6}$$

Suppose that $\phi_n(x)$ is an empirical Bayes estimator based on past data $(X_1, X_2, \cdots, X_n)$ and the present data $X = x$. Let Φ be the class of empirical Bayes estimators of type ϕ_n. We are interested in a lower bound of

$$\inf_{\phi_n \in \Phi} \sup_{G \in \mathcal{G}} [R(\phi_n, G) - R(\phi_G, G)]. \tag{7}$$

For any empirical Bayes estimator, Singh (1979) proved that

$$R(\phi_n, G) - R(\phi_G, G) = \int [E(\phi_n(x) - \phi_G(x))^2] \alpha_G(x) h(x) dx. \tag{8}$$

Therefore we have

$$\begin{aligned} & \inf_{\phi_n \in \Phi} \sup_{G \in \mathcal{G}} [R(\phi_n, G) - R(\phi_G, G)] \\ = & \inf_{\phi_n \in \Phi} \sup_{G \in \mathcal{G}} \int [E(\phi_n(x) - \phi_G(x))^2] \alpha_G(x) h(x) dx. \end{aligned} \tag{9}$$

The RHS of the above equation is a global minimax lower bound of empirical Bayes estimators instead of a local minimax lower bound of

$$\inf_{\phi_n \in \Phi} \sup_{G \in \mathcal{G}} [E(\phi_n(x) - \phi_G(x))^2] \quad \text{for some fixed } x. \tag{10}$$

So first we need to convert the global minimax problem into a local minimax problem. For this purpose, we focus on the supremum of the regret over two prior distributions and use the idea that the supremum of two positive numbers is larger than the half of the sum and the sum is larger than the supremum. Then we are able to move the "sup" into the integration in (9). By further moving the "inf" into the integration, a global minimax problem is changed into a point-wise (local) problem.

Lemma 2.1. *For any G_1, $G_2 \in \mathcal{G}$, let $\underline{\alpha}(x) = \alpha_{G_1}(x) \wedge \alpha_{G_2}(x)$ and*

$$\phi_{minimax}(x) = \inf_{\phi_n \in \Phi} \sup_{G \in \{G_1, G_2\}} [E(\phi_n(x) - \phi_G(x))^2]. \tag{11}$$

Then

$$\begin{aligned} & \inf_{\phi_n \in \Phi} \sup_{G \in \mathcal{G}} \int [E(\phi_n(x) - \phi_G(x))^2] \alpha_G(x) h(x) dx \\ \geq & \frac{1}{2} \int \phi_{minimax}(x) \underline{\alpha}(x) h(x) dx. \end{aligned} \tag{12}$$

Lemma 2.1 says that we can find $\phi_{minimax}(x)$ locally for each x and then obtain the global lower bound by integration. The proof of Lemma 2.1 is given in Section 5.

2.2 A Functional of the Marginal Density of X

Next we need to find $\phi_{minimax}(x)$ for each x. This is done by considering the estimate of a functional of the marginal density of X and constructing the hardest two-point subproblem associated with it.

Let $f_G(x) = \int f(x|\theta)dG(\theta)$ be the marginal density of X. Then $f_G(x) = \alpha_G(x)h(x)$. Assume that $h'(x)$ exists for $x \in (a,b)$. Then

$$\phi_G(x) = \frac{f'_G(x)}{f_G(x)} - \frac{h'(x)}{h(x)}. \tag{13}$$

For a fixed x, since $h(x)$ is known, the RHS of the above equation is a functional of $f_G(x)$. Let $T_x f_G$ denote this functional, i.e.,

$$T_x f_G \equiv \frac{f'_G(x)}{f_G(x)} - \frac{h'(x)}{h(x)} = \phi_G(x) \tag{14}$$

We have expressed the Bayes estimator $\phi_G(x)$ as a functional of f_G as above. To find ϕ_n is to estimate the functional $T_x f_G$ of f_G based on a sample from f_G. Therefore we apply the results in Donoho and Liu (1991) and obtain the following lemma.

Lemma 2.2. *Assume that $h'(x)$ exists for $x \in (a,b)$. For any G_1 and $G_2 \in \mathcal{G}$, let f_{G1} and f_{G2} be the corresponding marginal densities of X. If for some constant $C > 0$,*

$$\int [\sqrt{f_{G1}(x)} - \sqrt{f_{G2}(x)}]^2 dx \le \frac{C}{n}. \tag{15}$$

Then for all $x \in (a,b)$, $\phi_{minimax}(x)$ defined by (11) satisfies

$$\phi_{minimax}(x) \ge l_1[\phi_{G1}(x) - \phi_{G2}(x)]^2, \tag{16}$$

where $l_1 > 0$ is a constant and independent of x.

Combining Lemma 2.1 and Lemma 2.2, we have: for some $l_2 > 0$

$$\inf_{\phi_n \in \Phi} \sup_{G \in \mathcal{G}} [R(\phi_n, G) - R(\phi_G, G)] \tag{17}$$

$$\geq \; l_2 \int [\phi_{G1}(x) - \phi_{G2}(x)]^2 [\alpha_{G1}(x) \wedge \alpha_{G2}(x)] h(x) dx$$

for any G_1 and G_2 in $\mathcal{G}$ subject to

$$\int (\sqrt{f_{G1}(x)} - \sqrt{f_{G2}(x)})^2 dx \leq \frac{C}{n} \tag{18}$$

for some $C > 0$.

2.3 A Lower Bound

In the following we shall construct suitable G_1 and G_2 in $\mathcal{G}$ such that a desired lower bound of $\phi_{minimax}(x)$ can be obtained. Choose $x_0 \in (a, b)$. Let

$$g_0(\theta) = m_0 [c(\theta)]^{-1} I_{[\theta_1 \leq \theta \leq \theta_2]} \tag{19}$$

and

$$g_1(\theta) = m_1 \exp(\theta x_0) g_0(\theta) \tag{20}$$

where m_1 and m_2 are normalizing constants. Denote

$$g_2(\theta) = \frac{\sqrt{n} g_1(\theta) + g_0(\theta)}{1 + \sqrt{n}}. \tag{21}$$

Clearly, g_0, g_1 and g_2 are prior densities with their cdf's in $\mathcal{G}$.

Lemma 2.3. *Let g_1 and g_2 be defined as (20) and (21). Let f_1 and f_2 be the marginal densities of X corresponding to the prior density $g = g_1$ and $g = g_2$. Let $\alpha_0(x)$, $\alpha_1(x)$ and $\alpha_2(x)$ be the function $\alpha_G(x)$ corresponding to the prior density $g = g_0$, $g = g_1$ and $g = g_2$. Let $\psi_0(x)$, $\psi_1(x)$ and $\psi_2(x)$ be the function $\psi_G(x)$ corresponding to the prior density $g = g_0$, $g = g_1$ and $g = g_2$. Then for some constant $C > 0$*

$$\int (\sqrt{f_1} - \sqrt{f_2})^2 dx \leq \frac{C}{n} \tag{22}$$

and for all $x \in (a,b)$

$$(\phi_1(x) - \phi_2(x))^2 \geq \frac{l_3}{n} \times \frac{[\alpha_1(x)\psi_0(x) - \alpha_0(x)\psi_1(x)]^2}{\alpha_0^2(x)} \tag{23}$$

where l_3 is a constant independent of x.

The proof of Lemma 2.3 is given in Section 5. Under the assumption of Lemma 2.3, it follows from (17) and (18) that

$$\begin{aligned} & \inf_{\phi_n \in \Phi} \sup_{G \in \mathcal{G}} [(R(\phi_n, G) - R(\phi_G, G))^2] \\ \geq\ & \frac{l_2 l_3}{2n} \int \frac{[\alpha_0(x)\psi_1(x) - \alpha_1(x)\psi_0(x)]^2}{\alpha_0^2(x)} [\alpha_1(x) \wedge \alpha_2(x)] h(x) dx \\ \geq\ & \frac{l}{n} \end{aligned} \tag{24}$$

where

$$l = \frac{1}{2} l_2 l_3 \int_{x_1}^{x_2} \frac{[\alpha_0(x)\psi_1(x) - \alpha_1(x)\psi_0(x)]^2}{\alpha_0^2(x)} [\alpha_1(x) \wedge \alpha_2(x)] h(x) dx < \infty, \tag{25}$$

and $[x_1, x_2]$ is a compact subset of (a,b).

Theorem 2.1. *Assume that $h'(x)$ exists. Then the best possible rate of empirical Bayes estimators is $O(1/n)$, i.e., for some $l > 0$*

$$\inf_{\phi_n \in \Phi} \sup_{G \in \mathcal{G}} [R(\phi_n, G) - R(\phi_G, G)] \geq \frac{l}{n}. \tag{26}$$

Note that in the above theorem, we have proved that no empirical Bayes estimator can have a rate faster than $O(1/n)$ if θ is distributed within a compact (bounded) set. In the next section, we shall construct an empirical Bayes estimator with rate which is close to this minimax lower bound rate $O(1/n)$.

3 Construction of an Empirical Bayes Estimator with Rate Close to the Lower Bound Rate

In this section, we shall construct an empirical Bayes estimator under the assumption $\Omega \subset [\theta_1, \theta_2]$ and then show the estimator has a rate much closer to the best possible rate obtained in Section 2 than any other estimators appeared in the literature under the same assumptions.

Note that the Bayes estimator $\phi_G(x) = \psi_G(x)/\alpha_G(x)$, which is the ratio of two unknown functions. So we first estimate both unknown functions $\alpha_G(x)$ and $\psi_G(x)$. Then construct an estimator of a ratio based on the estimators of the numerator and denominator.

We apply the kernel sequence method to construct the estimators of $\alpha_G(x)$ and $\psi_G(x)$. The idea of the kernel sequence method is to use a sequence of kernel functions and let the kernel functions and window bandwidths vary simultaneously to obtain good estimators. This idea has been used in Gupta and Li (2001) for constructing an empirical Bayes test for the exponential family. It will be used here again.

3.1 Construction of an Estimator

We have defined two kernel sequences in Gupta and Li (2001) where we construct the empirical Bayes test for the exponential family. Unfortunately, they are not good choices for this estimation problem. So we have to define two different kernel sequences.

For $m \geq 1$, let

$$K_{0m}(y) = \begin{cases} p_m y^m + p_{m-1} y^{m-1} + \cdots + p_0, & \text{if } 0 \leq y \leq 1, \\ 0, & \text{otherwise,} \end{cases} \tag{27}$$

where for $0 \leq s \leq m$

$$p_s = \frac{(-1)^s (m+1)(m+s+1)!}{(s+1)!s!(m-s)!} \tag{28}$$

and let

$$K_{1m}(y) = \begin{cases} q_m y^m + q_{m-1} y^{m-1} + \cdots + q_0, & \text{if } 0 \le y \le 1, \\ 0, & \text{otherwise,} \end{cases} \tag{29}$$

where for $0 \le s \le m$

$$q_s = \frac{(-1)^{s+1}(m+2)!(s+m+1)!}{(m-1)!s!(s+2)s!(m-s)!}. \tag{30}$$

In Section 5 of this paper, we shall prove that

$$\int_0^1 y^j K_{0m}(y)dy = \begin{cases} 1 \text{ if } & j = 0, \\ 0 \text{ if } & j = 1, 2, \cdots, m, \end{cases} \tag{31}$$

and

$$\int_0^1 y^j K_{1m}(y)dy = \begin{cases} 0 \text{ if } & j = 0, 2, 3, \cdots, m, \\ 1 \text{ if } & j = 1. \end{cases} \tag{32}$$

So $k_{0m}(y)$ and $k_{1m}(y)$ are the kernels with index m. $K_{0m}(y)$ will be used to estimate $\alpha_G(x)$ and $K_{1m}(y)$ will be used to estimate $\psi_G(x)$. Clearly,

$$\begin{aligned} \int (K_{0m}(y))^2 dy &= \int (K_{0m}(y))(p_m y^m + p_{m-1} y^{m-1} + \cdots + p_0)dy \\ &= p_0 \end{aligned} \tag{33}$$

and

$$\int |K_{0m}(y)|dt \le \left(\int (k_{0m}(y))^2 dy\right)^{1/2} = p_0^{1/2}. \tag{34}$$

Similarly,

$$\int (K_{1m}(y))^2 dy = q_1 \tag{35}$$

and

$$\int |K_{1m}(y)|dy \le q_1^{1/2}. \tag{36}$$

Now we consider the following three cases of (1):

$$(*)\begin{cases} (I) & a=-\infty,\ \ b=\infty,\ \ h(x)\downarrow 0 \text{ as } x\downarrow -\infty \text{ and } x\uparrow\infty; \\ (II) & a=0,\ \ b=\infty,\ \ h(x)\uparrow\infty \text{ or } h(x)\to \underline{h}_1>0 \text{ as } x\uparrow\infty \text{ and} \\ & \qquad\qquad h(x)\to \underline{h}_2>0 \text{ or } h(x)\downarrow 0 \text{ as } x\downarrow 0; \\ (III) & a=0,\ \ b=1,\ \ h(x)\to \underline{h}_1>0 \text{ or } h(x)\downarrow 0 \text{ as } x\downarrow 0 \text{ and} \\ & \qquad\qquad h(x)\to \underline{h}_2>0 \text{ or } h(x)\downarrow 0 \text{ as } x\uparrow 1, \end{cases}$$

Note that $(*)$ includes the most common exponential family distributions.

Let $u=u_n=1/(\ln\ln n\vee 1)$, $v=v_n=[\frac{\ln n}{-\ln u}]\vee 0+1$, where $[x]$ denotes the integer part of x. In case I, define for $x\in(-\infty,0)$,

$$\begin{cases} \alpha_n(x)=\frac{1}{nu}\sum_{j=1}^n \frac{K_{0v}(\frac{X_j-x}{u})}{h(X_j)}, \\ \psi_n(x)=\frac{1}{nu^2}\sum_{j=1}^n \frac{K_{1v}(\frac{X_j-x}{u})}{h(X_j)}, \end{cases} \tag{37}$$

and for $x\in[0,\infty)$

$$\begin{cases} \alpha_n(x)=\frac{1}{nu}\sum_{j=1}^n \frac{K_{0v}(\frac{x-X_j}{u})}{h(X_j)}, \\ \psi_n(x)=\frac{1}{nu^2}\sum_{j=1}^n \frac{K_{1v}(\frac{x-X_j}{u})}{h(X_j)}, \end{cases} \tag{38}$$

In Case (II), define for $x\in(0,\infty)$

$$\begin{cases} \alpha_n(x)=\frac{1}{nu}\sum_{j=1}^n \frac{K_{0v}(\frac{X_j-x}{u})}{h(X_j)}, \\ \psi_n(x)=\frac{1}{nu^2}\sum_{j=1}^n \frac{K_{1v}(\frac{X_j-x}{u})}{h(X_j)}, \end{cases} \tag{39}$$

In Case (III), define for $x\in(0,1/2)$

$$\begin{cases} \alpha_n(x)=\frac{1}{nu}\sum_{j=1}^n \frac{K_{0v}(\frac{X_j-x}{u})}{h(X_j)}, \\ \psi_n(x)=\frac{1}{nu^2}\sum_{j=1}^n \frac{K_{1v}(\frac{X_j-x}{u})}{h(X_j)}, \end{cases} \tag{40}$$

and for $x \in [1/2, 1)$

$$\begin{cases} \alpha_n(x) = \frac{1}{nu}\sum_{j=1}^{n} \frac{K_{0v}(\frac{x-X_j}{u})}{h(X_j)}, \\ \\ \psi_n(x) = \frac{1}{nu^2}\sum_{j=1}^{n} \frac{K_{1v}(\frac{x-X_j}{u})}{h(X_j)}, \end{cases} \tag{41}$$

Lemma 3.1 below says that $\alpha_n(x)$ and $\psi_n(x)$ are consistent estimators of $\alpha_G(x)$ and $\psi_G(x)$.

Note that $\theta_1 \le \theta \le \theta_2$. Therefore we propose an empirical Bayes estimator of θ as

$$\phi_n(x) = (\frac{\psi_n(x)}{\alpha_n(x)} \vee \theta_1) \wedge \theta_2. \tag{42}$$

3.2 Rate of Convergence of the Estimator

First we investigate the rate of convergence of α_n and ψ_n. For the distributions of Case I, Case II and Case III, we see that $h(x)$ is either bounded from below or monotone in $(a, a_0]$ and $[b_0, b)$ for some a_0 and b_0. Let

$$\underline{h}(x) = h(x) \wedge \min\{h(x) : x \in [a_0, b_0]\}. \tag{43}$$

As a result of the kernel sequence estimation, we have the following lemma.

Lemma 3.1. *$\alpha_n(x)$ and $\psi_n(x)$ defined in (37)-(41) have the following properties:*

$$|E[\alpha_n(x)] - \alpha_G(x)| \le c_1 p_0^{1/2} u^v \alpha_G(x), \quad Var[\alpha_n(x)] \le c_2 \frac{p_0 \alpha_G(x)}{nu\underline{h}(x)}, \tag{44}$$

and

$$|E[\psi_n(x)] - \psi_G(x)| \le c_1 q_1^{1/2} u^v \alpha_G(x), \quad Var[\psi_n(x)] \le c_2 \frac{q_1 \alpha_G(x)}{nu\underline{h}(x)}. \tag{45}$$

where c_1, c_2 are constants and independent of x and G.

From Lemma 3.1, we see that the mean square errors of $\alpha_n(x)$ and $\psi_n(x)$ are of order $O(\frac{1}{nu})$ or $O(\frac{\ln\ln n}{n})$. This fast rate is due to the use of the kernel sequence method.

The following two lemmas are necessary to compute the convergence rate of ϕ_n. The first one gives a bound on the mean squared error of ϕ_n.

Lemma 3.2. *For any* $0 < r \leq 1$

$$
\begin{aligned}
E[|\phi_n(x) - \phi_G(x)|^2] \quad &\leq \quad c_3\alpha_G^{-2r}(x)[\{|E[\alpha_n(x)] - \alpha_G(x)|\}^{2r} + \{Var[\alpha_n(x)]\}^r] \\
&\quad + c_4\alpha_G^{-2r}(x)[\{|E[\psi_n(x)] - \psi_G(x)|\}^{2r} + \{Var[\psi_n(x)]\}^r],
\end{aligned}
$$

where c_3, c_4 *are constants and independent of* x, r *and* G.

Lemma 3.3. *Recall* $f_G(x) = \int f(x|\theta)dG(\theta)$. *For any* $0 < r < 1$

$$
\int [f_G(x)]^{1-r}dx \leq \begin{cases} c_5/(1-r) & \textit{for Case I and II}, \\ c_5 & \textit{for Case III}, \end{cases} \tag{46}
$$

where c_5 *is a constant and independent of* r *and* G.

The proofs of Lemma 3.1 and Lemma 3.3 are in Section 5 and Lemma 3.2 is a modified version of Lemma 2.1 in Singh (1979). From Lemma 3.1 and 3.2, we have

$$
\begin{aligned}
R(\phi_n, G) - R(\phi_G, G) \quad &= \quad \int E[|\phi_n(x) - \phi_G(x)|^2]\alpha_G(x)h(x)dx \qquad (47) \\
&\leq \quad c_3(c_1p_0^{1/2}u^v)^{2r} + c_3\{c_2p_0(nu)^{-1}\}^r \int \alpha_G^{1-r}(x)h(x)\underline{h}^{-r}(x)dx \\
&\quad + c_4(c_1q_1^{1/2}u^v)^{2r} + c_4\{c_2q_1(nu^3)^{-1}\}^r \int \alpha_G^{1-r}(x)h(x)\underline{h}^{-r}(x)dx
\end{aligned}
$$

Note that $\underline{h}(x) \geq c_6h(x)$ for some constant $c_6 > 0$. Then

$$
\int \alpha_G^{1-r}(x)h(x)\underline{h}^{-r}(x)dx \leq c_6^{-1}\int [\alpha_G(x)h(x)]^{1-r}dx = c_6^{-1}\int [f_G(x)]^{1-r}dx. \tag{48}
$$

Also note that $u^v \leq n^{-1}$, $p_0 = (v+1)^2$ and $q_1 = [v(v+1)(v+2)]^2/3$. Then for Case I and II

$$
R(\phi_n, G) - R(\phi_G, G) = c_7 \cdot \frac{(v+1)^{6r}}{(nu^3)^r} \cdot \frac{1}{1-r} \leq c_7\frac{(\ln n)^7}{n}(\ln\ln n)^2, \tag{49}
$$

where $c_7 = (c_3 + c_4)(c_1 + c_2)c_5/(3c_6)$ and $r = 1 - \ln\ln n/\ln n$, and for Case III

$$
R(\phi_n, G) - R(\phi_G, G) = c_7 \cdot \frac{(v+1)^6}{nu^3} \leq c_7\frac{(\ln n)^6}{n}(\ln\ln n)^3. \tag{50}
$$

Then we have the following theorem.

Theorem 3.1. *Let the distribution of X belong to one of three cases defined by $(*)$. If the support of the prior G is within a compact set $[\theta_1, \theta_2] \subset \Omega_0$, then the empirical Bayes estimator $\phi_n(x)$ defined by (42) has a rate of convergence of $O(n^{-1}(\ln n)^7(\ln\ln n)^2)$ for the distributions in Case I and Case II, and has a rate of convergence of $O(n^{-1}(\ln n)^6(\ln\ln n)^3)$ for the distributions in Case III.*

Corollary 3.1. *Under the assumption of Theorem 3.1 and Let $\mathcal{G}$ be the set of prior distributions defined by (6). Then for the distributions in Case I and II*

$$\sup_{G\in\mathcal{G}}[R(\phi_n, G) - R(\phi_G, G)] = O(n^{-1}(\ln n)^7(\ln\ln n)^2) \tag{51}$$

and for the distributions in Case III

$$\sup_{G\in\mathcal{G}}[R(\phi_n, G) - R(\phi_G, G)] = O(n^{-1}(\ln n)^6(\ln\ln n)^3). \tag{52}$$

From Theorem 2.1, we know that $O(1/n)$ is the best possible rate. From Corollary 3.1, we see that ϕ_n constructed by (42) has a rate close to $O(1/n)$. Comparing the previous results in the literature, the rate in Corollary 3.1 is the fastest one under the same assumptions. See Section 6 for details on comparisons.

4 Examples

We shall present a few examples in this section. The first three are from Singh (1979). Another example is used to illustrate the application of the empirical Bayes rule ϕ_n for the distribution in Case III. So a brief comparison of our results with the results published in the literature will be presented. For a more comprehensive comparison, see Section 6.

Example 1. (*Normal* $(\theta, 1)$*-family*). Suppose that X is a normal random variable with density

$$f(x|\theta) = (2\pi)^{-1/2} \exp(-\theta^2/2) \exp(\theta x) \exp(-x^2/2), \quad -\infty < x < \infty.$$

Here the natural parameter space $\Omega_0 = (-\infty, \infty)$. If θ is bounded, i.e. if $|\theta| \leq \theta_0$, then ϕ_n of (42) with $\theta_1 = -\theta_0$ and $\theta_2 = \theta_0$ has a rate of convergence of $O(n^{-1}(\ln n)^7(\ln\ln n)^2)$. Note that the rate of Singh's estimator is close to $O(n^{-2(r-1)/(1+2r)})$ for $r > 1$. So under the same assumption, our rate is faster.

Example 2. (*Gamma* $(\theta,\ s)$*-family for* $s > 1$). Suppose that X is a gamma random variable with density

$$f(x|\theta) = (\Gamma(s))^{-1}(-\theta)^{-s} \exp(\theta x) x^{s-1}, \qquad x > 0, \quad s > 1.$$

Here the natural parameter space $\Omega_0 = (-\infty, 0)$. If $-\infty < \theta_1 \leq \theta \leq \theta_2 < 0$, then ϕ_n of (42) has a rate of convergence of $O(n^{-1}(\ln n)^7(\ln\ln n)^2)$, which is better than Singh's polynomial rate in Singh (1979).

Example 3. (*A population having the density with infinite many discontinuities*). Suppose that X is a random variable with density

$$f(x|\theta) = (-\theta)(\exp(\theta) - 1) \exp(\theta x) \sum_{l=0}^{\infty} (l+1) I_{[l < x \leq l+1]}, \qquad x > 0.$$

Here the natural parameter space $\Omega_0 = (-\infty, 0)$. For this distribution, Theorem 3.1 is applicable and our rate $O(n^{-1}(\ln n)^7(\ln\ln n)^2)$ is better than Singh's rate $O(n^{-2(r-1)/(1+2r)})$ under the assumption that $\Omega \in [\theta_1, \theta_2] \subset \Omega_0$. Since $h(x)$ is not differentiable, Pensky's method in Pensky (1998) fails to giving the rate of convergence.

Now we give one example for the application of Theorem 3.1 in Case III distributions.

Example 4. Suppose that X is a random variable from the following truncated exponential distribution:

$$f(x|\theta) = \theta(1-\exp(\theta))^{-1}\exp(\theta x), \qquad 0 < x < 1.$$

Here the natural parameter space $\Omega = (0, \infty)$. If $0 < \theta_1 \le \theta \le \theta_2 < \infty$, then $\phi_n(x)$ of (42) has a rate of of convergence $O(n^{-1}(\ln n)^6(\ln\ln n)^3)$.

5 Proofs.

Proof of Lemma 2.1. For any $G_1, G_2 \in \mathcal{G}$, we have

$$\inf_{\phi_n\in\Phi}\sup_{G\in\mathcal{G}}[R(\phi_n, G) - R(\phi_G, G)] \ge \inf_{\phi_n\in\Phi}\sup_{G\in\{G_1,G_2\}}[R(\phi_n, G) - R(\phi_G, G)] \tag{53}$$

Then it follows that

$$\begin{aligned}
&\inf_{\phi_n\in\Phi}\sup_{G\in\mathcal{G}}\int[E(\phi_n(x)-\phi_G(x))^2]\alpha_G(x)h(x)dx \qquad (54)\\
\ge\ & \frac{1}{2}\inf_{\phi_n\in\Phi}\Big[\int[E(\phi_n(x)-\phi_{G1}(x))^2]\alpha_{G1}(x)h(x)dx \\
&\qquad + \int[E(\phi_n(x)-\phi_{G2}(x))^2]\alpha_{G2}(x)h(x)dx\Big] \\
\ge\ & \frac{1}{2}\inf_{\phi_n\in\Phi}\Big[\int[E(\phi_n(x)-\phi_{G1}(x))^2]\underline{\alpha}(x)h(x)dx \\
&\qquad + \int[E(\phi_n(x)-\phi_{G2}(x))^2]\underline{\alpha}(x)h(x)dx\Big] \\
\ge\ & \frac{1}{2}\inf_{\phi_n\in\Phi}\int\sup_{G\in\{G_1,G_2\}}[E(\phi_n(x)-\phi_{G1}(x))^2]\underline{\alpha}(x)h(x)dx \\
\ge\ & \frac{1}{2}\int\inf_{\phi_n\in\Phi}\sup_{G\in\{G_1,G_2\}}[E(\phi_n(x)-\phi_{G1}(x))^2]\underline{\alpha}(x)h(x)dx
\end{aligned}$$

This completes the proof of Lemma 2.1.

Proof of Lemma 2.3. From (21), it is clear that

$$f_2(x) - f_1(x) = \frac{-f_1(x)+f_0(x)}{1+\sqrt{n}} \tag{55}$$

Note that $f_0(x) = m_0\int_{\theta_1}^{\theta_2}\exp(\theta x)d\theta$ and $f_1(x) = m_1 m_0\int_{\theta_1}^{\theta_2}\exp(\theta(x+x_0))d\theta$. Then there exist $l_1 > 0$ and $l_2 > 0$ such that for all $x \in (a, b)$,

$$l_1 \leq \frac{f_0(x)}{f_1(x)} = \frac{\int_{\theta_1}^{\theta_2} \exp(\theta x) d\theta}{\int_{\theta_1}^{\theta_2} \exp(\theta(x + x_0)) d\theta} \leq l_2 \tag{56}$$

Therefore

$$\begin{aligned} \int [\sqrt{f_1(x)} - \sqrt{f_2(x)}]^2 dx &\leq \int \frac{[f_1(x) - f_2(x)]^2}{f_1(x)} dx \\ &\leq \frac{1}{(1+\sqrt{n})^2} \int [1 - \frac{f_0(x)}{f_1(x)}]^2 f_1(x) dx \\ &\leq \frac{C}{n} \end{aligned} \tag{57}$$

for some $C > 0$. On the other hand,

$$\begin{aligned} \phi_2(x) - \phi_1(x) &= \frac{\sqrt{n}\psi_1(x) + \psi_0(x)}{\sqrt{n}\alpha_1(x) + \alpha_0(x)} - \frac{\psi_1(x)}{\alpha_1(x)} \\ &= \frac{\psi_0(x)\alpha_1(x) - \alpha_0(x)\psi_1(x)}{\alpha_1(x)[\sqrt{n}\alpha_1(x) + \alpha_0]} \end{aligned} \tag{58}$$

Since

$$l_1 \leq \frac{\alpha_0(x)}{\alpha_1(x)} = \frac{f_0(x)}{f_1(x)} \leq l_2 \tag{59}$$

There exists $l > 0$ such that

$$(\phi_2(x) - \phi_1(x))^2 \geq \frac{l}{n} \times \frac{[\alpha_1(x)\psi_0(x) - \alpha_0(x)\psi_1(x)]^2}{\alpha_0^2(x)} \tag{60}$$

This completes the proof of Lemma 2.3.

Proof of (31) and (32). To prove (31), it is sufficient to show that

$$\begin{cases} p_0 + \frac{p_1}{2} + \cdots + \frac{p_m}{m+1} & = 1 \\ \frac{p_0}{2} + \frac{p_1}{3} + \cdots + \frac{p_m}{m+2} & = 0 \\ \cdots \quad \cdots \quad \cdots & \\ \frac{p_0}{m+1} + \frac{p_1}{m+2} + \cdots + \frac{p_m}{2m+1} & = 0. \end{cases} \tag{61}$$

Then we need to show that p_s $(0 \leq s \leq m)$ is the solution of (61). Using Cramer's rule, we have, for $0 \le s \le m$,

$$p_s = \frac{\begin{vmatrix} 1 & \cdots & 1 & \cdots & \frac{1}{m+1} \\ \frac{1}{2} & \cdots & 0 & \cdots & \frac{1}{m+2} \\ \cdots & \cdots & \cdots & \cdots & \\ \frac{1}{m+1} & \cdots & 0 & \cdots & \frac{1}{2m+1} \end{vmatrix}}{\begin{vmatrix} 1 & \cdots & \frac{1}{s+1} & \cdots & \frac{1}{m+1} \\ \frac{1}{2} & \cdots & \frac{1}{s+2} & \cdots & \frac{1}{m+2} \\ \cdots & \cdots & \cdots & \cdots & \\ \frac{1}{m+1} & \cdots & \frac{1}{s+m+1} & \cdots & \frac{1}{2m+1} \end{vmatrix}} = \frac{det_2}{det_1},$$

where det_2 is the numerator of p_s and det_1 is the denominator of p_s. A simple calculation shows that

$$det_1 = \frac{[m!(m-1)!\cdots 2!]^3}{(m+1)!(m+2)!\cdots(2m+1)!}$$

and

$$det_2 = \frac{(-1)^{s+2}(m!)^2[(m-1)!(m-2)!\cdots 2!]^3(s+m+1)!}{(m+2)!(m+3)!\cdots(2m+1)!(s+1)!s!(m-s)!}.$$

Thus

$$p_s = \frac{(-1)^s(m+1)(m+s+1)!}{(s+1)!s!(m-s)!}.$$

So (31) is proved. The proof of (32) is similar. It is omitted here.

Proof of Lemma 3.1. We prove (44) only. The proof of (45) is similar. Let $\theta_0 = |\theta_1| \vee |\theta_2|$. In Case I, using Taylor expansion and (31).

$$\begin{aligned} & E[\frac{K_{0v}(\frac{X_j-x}{u})}{uh(X_j)}] \\ = & \int_\Omega c(\theta)e^{\theta x}dG(\theta) + u^v \int_\Omega c(\theta)e^{\theta x}[\int_0^1 K_{0v}(t)e^{\theta u t^*}\frac{(\theta t)^v}{v!}dt]dG(\theta), \end{aligned} \tag{62}$$

where $t^* \in (0,1)$. Note that

$$\int_0^1 K_{0v}(t)e^{\theta u t^*}\frac{(\theta t)^v}{v!}dt \le e^{2\theta_0}\int_0^1 |K_{0v}(t)|dt = e^{2\theta_0}p_0^{1/2}. \tag{63}$$

Then

$$|E[\alpha_n(x)] - \alpha_G(x)| \le e^{2\theta_0} p_0^{1/2} u^v \alpha_G(x) \tag{64}$$

Note that for $x < 0$, $0 < t < 1$, $h(x+ut) \ge \underline{h}(x)$ and

$$Var[\alpha_n(x)] \le \frac{1}{nu} \int \frac{[K_{0v}(t)]^2}{h(x+ut)} c(\theta) e^{\theta x + \theta u t} dt dG(\theta) \le \frac{e^{\theta_0}}{nu\underline{h}(x)} p_0 \alpha_G(x). \tag{65}$$

Similarly, $Var[\alpha_n(x)] \le e^{\theta_0} p_0 \alpha_G(x)/[nu\underline{h}(x)]$ for $x > 0$. Then (44) is proved.

Proof of Lemma 3.3. We prove (46) for different cases of $(*)$.

Case I. Let η satisfy $\theta_1 - \eta \in \Omega$ and $\theta_2 + \eta \in \Omega$. For any $\theta \in [\theta_1 - \eta, \theta_2 + \eta]$, $f(x|\theta)$ is bounded on $(-\infty, \infty)$. For any $(\theta, x) \in [\theta_1 - \eta, \theta_2 + \eta] \times [0, \infty)$, we have

$$\begin{aligned} f(x|\theta) &= c(\theta) \exp(\theta x) h(x) \\ &\le c(\theta) \exp((\theta_2 + \eta)x) h(x) \\ &= \frac{c(\theta)}{c(\theta_2 + \eta)} f(x|\theta = \theta_2 + \eta). \end{aligned} \tag{66}$$

For any $(\theta, x) \in [\theta_1 - \eta, \theta_2 + \eta] \times (-\infty, 0)$, we have

$$f(x|\theta) \le \frac{c(\theta)}{c(\theta_1 - \eta)} f(x|\theta = \theta_1 - \eta). \tag{67}$$

Since $c(\theta)$ is a convex function on $[\theta_1 - \eta, \theta_2 + \eta]$, it follows from (66) and (67) that there exists $M > 1$ such that for any $(\theta, x) \in [\theta_1 - \eta, \theta_2 + \eta] \times (-\infty, \infty)$

$$f(x|\theta) \le M. \tag{68}$$

Let $c = \max_{\theta \in [\theta_1, \theta_2]} \{ \frac{c(\theta)}{c(\theta+\eta)}, \frac{c(\theta)}{c(\theta-\eta)} \} \vee 1$. Then $c < \infty$. And

$$\begin{aligned} \int [f_G(x)]^{1-r} dx &= \int_0^\infty [\int f(x|\theta) dG(\theta)]^{1-r} dx + \int_{-\infty}^0 [\int f(x|\theta) dG(\theta)]^{1-r} dx \\ &\le c \int_0^\infty [\int c(\theta + \eta) \exp((\theta + \eta)x) h(x) dG(\theta)]^{1-r} \exp(-\eta(1-r)x) dx \\ &\quad + c \int_{-\infty}^0 [\int c(\theta - \eta) \exp((\theta - \eta)x) h(x) dG(\theta)]^{1-r} \exp(\eta(1-r)x) dx \end{aligned} \tag{69}$$

$$
\begin{aligned}
&\leq cM[\int_0^\infty \exp(-\eta(1-r)x)dx + \int_{-\infty}^0 \exp(\eta(1-r)x)dx] \\
&= \frac{2cM}{\eta(1-r)}
\end{aligned}
$$

Case III. Note that there exists $M > 1$ such that for any $\theta \in [\theta_1, \theta_2]$ and for any $x \in (0,1)$

$$f(x|\theta) \leq M. \tag{70}$$

Then

$$\int_0^1 [f_G(x)]^{1-r} dx = \int_0^1 M^{1-r} dx \leq M. \tag{71}$$

Case II. Note that

$$\int_0^\infty [f_G(x)]^{1-r} dx = \int_0^1 [f_G(x)]^{1-r} dx + \int_1^\infty [f_G(x)]^{1-r} dx. \tag{72}$$

Then Lemma 3.3 in this case follows the methods used in the proofs for Case I and Case III.

6 Summary and Discussion

In this paper, we have studied the estimation problem in the exponential family. First we proved that the best possible rate of empirical Bayes estimators is $O(1/n)$ if θ is distributed within a bounded compact set. This gives a goal that we are working toward in constructing the empirical Bayes estimators. For a long time, people have been thinking that $O(1/n)$ is a natural lower bound rate. But it had never been proved.

Also we have constructed an estimator which achieves a rate of $O(n^{-1}(\ln n)^7(\ln\ln n)^2)$ under the assumption that θ is distributed within a bounded compact set. Under the same assumption, this is the fastest rate comparing to the rates that have appeared in the literature before.

Most recent significant results on this estimation problem are published by Singh (1979) and Pensky (1998). In their papers, they constructed the empirical Bayes estimators and investigated the convergence rate of the estimators. Singh (1979) used the kernel method and Pensky (1998) applied the advanced wavelet techniques in their construction. Both papers allow the unboundedness of θ but get a polynomial rate. For Singh's result, the rate will stay the same even under additional assumption that Ω is a compact set. So our result is much better than his under the same assumption. To get a rate like we have here from Pensky's result, the existence of all moments of θ is necessary. Also the degree of smoothness of $f(x|\theta)$ is a key factor to determine the rate of convergence. If the degree of smoothness is low, the rate is slow even if θ is distributed within a compact set.

References

[1] Bender, R. B. and Carlin, B. P. (1990). Bayes methods in the ecological fallacy context: estimation of individual correlation from aggregate data. *Commun. Statist. Theory Methods*, **19**, 2595-2623.

[2] Chen, J. and Singpurwalla, N. D. (1996). The notion of " composite reliability" and its hiearchical Bayes estimation. *J. Amer. Statist. Assoc.*, **436**, 1474-1484.

[3] Desouza, C. M. (1991). An empirical Bayes formulation of cohort models in cancer epidemilogy. *Stat. Med.*, **10**, 1241-1256.

[4] Donoho, D. L. and Liu, R. C. (1991). Geometrizing rates of convergence. II. *Ann. Statist.* **19**, 633-667.

[5] Gupta, S. S. and Li, J. (2000). Monotone empirical Bayes tests based on kernel sequence estimation.

[6] Lahiri, P. and Park, D. H. (1991). Nonparametric Bayes and empirical Bayes estimators of mean residual life at age *t*. *JSPI*, **29**, 125-136.

[7] Louis, T. A. (1991). Using empirical Baues methods in biopharmaceutical research. *Stat. Med.*, **10**, 811-827.

[8] Mollie, A. and Richardson, S. (1991). Empirical Bayes estimates of cancer mortality rates using spatial models. *Stat. Med.*, **10**, 95-112.

[9] Pensky, M. (1998). Empirical Bayes estiamtion based on wavelets. *Sankya*, Series A, **60**, 214-231.

[10] Pensky, M. and Singh R. S. (1999). Empirical Bayes estiamtion of reliability characteristics for an exponential family. *Canadian J. of Stat.*, **27**, 127-136.

[11] Robbins, H.(1956). An empirical Bayes approach to statistics. *Proc. Third Berkeley Symp. Math. Statist. Probab.* **1**, 157-163, University of California Press, Berkeley.

[12] Robbins, H. (1964). The empirical Bayes approach to statistical decision problems. *Ann. Math. Statist.* **35**, 1-20.

[13] Singh, R. S. (1976). Empirical Bayes estimation with convergence rates in non-continuous Lebesgue-exponential families. *Ann. Statist.* **4**, 431-439.

[14] Singh, R. S. (1979). Empirical Bayes estimation in Lebesgue-exponential family with rates near the best possible rate. *Ann. Statist.* **7**, 890-902.